HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2013 January 01.

Published in final edited form as:

Nat Methods. ; 9(7): 743–748. doi:10.1038/nmeth.2069.

Single cell systems biology by super-resolution imaging and combinatorial labeling

Eric Lubeck[1] and **Long Cai**[1,2]

[1]Program in Biochemistry and Molecular Biophysics

[2]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena CA 91125 USA

Abstract

Fluorescence microscopy is a powerful quantitative tool for exploring regulatory networks in single cells. However, the number of molecular species that can be measured simultaneously is limited by the spectral separability of fluorophores. Here we demonstrate a simple but general strategy to drastically increase the capacity for multiplex detection of molecules in single cells by using optical super-resolution microscopy (SRM) and combinatorial labeling. As a proof of principle, we labeled mRNAs with unique combinations of fluorophores using Fluorescence *in situ* Hybridization (FISH), and resolved the sequences and combinations of fluorophores with SRM. We measured the mRNA levels of 32 genes simultaneously in single *S. cerevisiae* cells. These experiments demonstrate that combinatorial labeling and super-resolution imaging of single cells provides a natural approach to bring systems biology into single cells.

Introduction

Systems biology seeks to quantitatively understand the interactions between many biological components. To do so two distinct technical approaches are employed: genomics[1-4], which permits the analysis of all genes and proteins in a system simultaneously, and single cell biology[5-10], which follows the behavior of genes in individual cells within their native spatial context. However, these approaches have complementary limitations: genomics averages over heterogeneity and spatial complexity within populations, while single-cell methods have been limited to the study of a few genes at a time.

To unify the two approaches, we propose to use super-resolution microscopy to bring genomics into single cells. SRM[11-14] has been a powerful tool for cell biology and can resolve sub-cellular structures down to 10-20 nm. This extraordinary resolution can be harnessed for systems biology: for the typical yeast cell of 100 μm^3, the 10-20 nm resolution

Correspondences should be sent to lcai@caltech.edu.

Author Contributions

E.L. and L.C. performed the experiments, carried out the analysis, and wrote the manuscript. L.C. conceived the idea and designed the experiments.

The authors declare there are no competing financial interest.

of SRM translates into approximately 10^8 independent volume elements (voxels) per cell. These voxels provide enough space to detect large numbers of fluorophore-based barcodes attached to molecular species. Using only 9 of the photoswitchable fluorophore pairs currently available for SRM, over a hundred combinations of quadruplet ($_9C_4$=126) barcodes can be used. The abundances of each molecule can be quantified by counting the number of times the corresponding barcode is observed in the super-resolution image of the cell, within their native cellular and inter-cellular contexts.

To demonstrate the feasibility of this approach, we performed proof-of-principle experiments to detect multiple mRNA species in single *Saccharomyces cerevisiae* cells. We based our approach on the single molecule FISH (smFISH) technique[15,16]. We used smFISH to combinatorially barcode transcripts, taking advantage of the high labeling specificity of oligo probes. We employed two different barcoding strategies, spatial and spectral. The first relied on the spatial ordering of fluorophores from many small oligonucleotide probes bound to the transcripts. The second relied only on the combination of colors used in the probes. Although the first strategy allows higher multiplexing, the second strategy is less technically demanding and more robust. We therefore used spectral barcoding to profile transcripts from 32 stress responsive genes in single *S. cerevisiae* cells to demonstrate the method. In comparison to previous multiplex FISH approaches that labeled chromosomal loci and transcriptional active sites[17-18] our approach directly barcoded single mRNAs.

Results

Spatial coding of single mRNAs

The first strategy we explored for combinatorial labeling directly resolved the spatial ordering of different fluorescent oligonucleotide probes on individual mRNAs. As 20mer probes are approximately 7 nm in length, regions of the mRNA separated by 100 nucleotides and hybridized as a group of 4-5 probes can in principle be resolved by SRM with its 10-20 nm resolution. By labeling the probes with different fluorophores that hybridize to mRNA in a specific pattern, a nanoscopic barcode can be resolved on each transcript (Fig. 1).

We targeted the *PUN1* mRNA in single *S. cerevisiae* cells with 3 sets of oligo probes labeled with different fluorophores. These probes were tiled along the mRNA in a 5′ to 3′ spatially ordered fashion. Hybridized mRNAs appeared as co-localized and diffraction-limited spots (Fig. 1a). We observed that 96±2% (*n*=29 molecules, s.e.m.) of spots co-localized in all three channels, indicating efficient hybridization of the *PUN1* probes. We quantified the hybridization efficiency of a single FISH probe to be H_{single} =67.5±9.1**%** by photobleaching (Supplementary Fig.1, *n*=173, s.e.m.). The effective hybridization efficiency of the barcode, or the probability that at least one probe is hybridized for each coding position is $H_{barcode}=(1-(1-H_{single})^R)^N$, where *R* is the probe redundancy at each coding position and *N* is the number of coding positions. With 4 probes in each code position, the labeling efficiency is 99% for each position giving a 96% coincidence rate for all 3 coding positions.

The position of labeled probes on mRNA can be determined to a much higher resolution by gaussian fitting[19,20] of the co-localized spots. We observed the correct spatial order in 74±8% (*n*=28, s.e.m.) of 3-color co-localized mRNAs following alignment. Error in detecting the order of the barcode (26%) may result from a combination of factors, including localization error, lack of z resolution, and mRNA secondary structure (Supplementary Note). Barcode readout fidelity was robust to switching the barcode order (Fig. 1e-h), independent of mRNA identity (Supplementary Fig. 2-4)

Super-resolution imaging of mRNA barcodes

Barcoding with conventional fluorescence microscopy is only useful in cases where transcript levels are low. When the density of transcripts is high, diffraction-limited fluorescent spots of the same color will overlap and make barcode readout impossible. Super-resolution imaging using sparse activation of subsets of overlapping fluorophores is essential for high-density multiplex barcoding when signals from multiple transcripts overlap.

To perform super-resolution barcoding, we turned to STORM (Stochastic Optical Reconstruction Microscopy) imaging with Cyanine dye based photoswitchable dye pairs[21]. We used singly-labeled probes that only generated photoswitchable pairs when an emitter-labeled and an activator-labeled oligonucleotide probe hybridized next to one another on the mRNA (Fig. 2a). As both probes are required for the fluorophore to be re-activated, background is dramatically reduced because non-specifically bound emitter probes cannot reactivate. We used the activator-emitter dye pairs to construct our barcodes, with Cy5, Alexa 680 and Alexa 750 as emitters (Supplementary Fig. 5). and Alexa 405, FITC and Cy3 as activators[21].

We labeled the *PUN1* mRNA (Fig. 2a) using Cy5-based probe pairs. With 4 probe-pair redundancy, we observed a 3 color hybridization rate of 61±8% (*n*=85, s.e.m.), consistent with the theoretical hybridization efficiency of $H_{superbar}$=75% (Supplementary Note and Fig. 1). Hybridization efficiency can be improved by using a higher redundancy in coding: For example with 6 probe redundancy, the hybridization rate increase to 92% for a 3 position barcode. Of the barcodes that were completely hybridized, 72±10% (*n*=50, s.e.m.) were reconstructed correctly, consistent with the conventional fluorophore results and independent of mRNA species (Fig. 2 a,b, Supplementary Figs. 6-8).

For two-dimensional imaging the spatial barcoding scheme necessitates the linearization of mRNA. To do so, we compressed cells between coverglass slips, extending the mRNA through the lateral force of compression. Such treatment can be readily applied to single cell organisms and embryos, but may destroy spatially complex samples (Supplementary Note).

The principle drawback of spatial barcoding is its high-resolution requirement. It is difficult to use dyes that emit lower numbers of photons, such as Alexa 680 and Alexa 750, when high (~20 nm) intramolecular resolution is necessary. We anticipate that the continued development of improved fluorophores, 3D microscopy, and nucleic acid self-assembly[22] will enable spatial coding to reach its full potential. In the meantime an alternative strategy which we call spectral barcoding is more appropriate for most applications.

Spectral Coding

In spectral barcoding (Fig. 2 e, f) the identity of mRNA is coded by the combination of fluorophores, ignoring spatial order. As long as all the colors are present (as determined by the integrated photon counts over the STORM imaging cycle) and can be identified above the cross-talk tolerances, the barcode assignment can be confidently made even if the total amount of collected photons is low. This comparatively low resolution requirement enables the use of dim fluorophores, as probes need only to be localized to within the ~100 nm area of a single mRNA (Table 1). Probes labeled with the same fluorophore can be distributed throughout an mRNA, making spectral coding more robust to heterogeneities in hybridization and partial degradation of mRNA. In contrast to spatial coding, molecules do not need to be linearized to be faithfully identified, potentially permitting other molecules, such as proteins, to be multiplexed under super-resolution imaging.

Profiling stress response genes in single yeast cells

As a proof of principle of our technique, we profiled transcripts from 32 known stress responsive genes (Supplementary Table 1) in single *S. cerevisiae* cells in response to extracellular Calcium stress. We have previously shown that the master transcription factor Crz1 translocates in and out of the nucleus in short (2-3 minute) well-defined pulses in the presence of extracellular Calcium[23]. These pulses occur stochastically in time and activate target gene expression in a probabilistic fashion. In addition to Crz1, calcium stress also triggers the Msn2 pathway, a general stress response regulator that also pulses in its localization[23].

For mRNA profiling we selected 14 genes that are regulated by Crz1, 5 general stress response genes, as well as 13 other aging and stress markers[24]. To ensure that we were observing the products of individual TF pulses, we fixed cells under conditions where the average interval between pulses[23] was longer than the typical mRNA lifetime.

To label these genes, we used spectral coding of combinations of 3 out of 7 photoswitchable dye pairs ($_3C_7$=35). Barcode assignments are shown in Supplementary Table 1. Cells were imaged by STORM and transcript levels in individual cells quantified based on the abundances of the corresponding barcodes (Supplementary Fig. 9). The main concern with multiplexing a large probe set is the crosstalk among the barcodes. This can arise from several sources: fluorescent background from the cell, blinking of nonspecifically bound probes, crosstalk amongst the fluorophores, and errors in computational analysis. To rule out crosstalk and non-specific blinking, we only accepted fluorophore assignments at least 3 standard deviations from the measured crosstalk values (Supplementary Fig. 10). To rule out the possibility that random autofluoresence from cells would skew the measurements, we implemented an imaging scheme containing frequent frames without activation to reject high-frequency false-activation (Supplementary Fig. 11). In addition, we intentionally left certain barcodes unassigned and measured the rate of false positives from those barcodes to be 0.67±0.84 (n=62 cells, s.d.) copies per cell, which represents the additive noise floor of our measurements.

To check the accuracy of our quantitation by super-resolution barcoding, we compared the mean expressions level with qPCR as well as single molecule FISH measurements, and obtained R^2=0.95 for both experiments (Fig. 3a, b). We found a 2 fold difference between mean copy number measured with smFISH vs super-resolution barcodes, which can be accounted for by the hybridization efficiency of our probe set and the high threshold levels used to reject false activations. For future probe set designs, a higher redundancy in probes will considerably increase the effective hybridization rate.

To determine if there is an appreciable bias introduced by a particular barcode scheme, we compared the expression levels measured by 2 different scrambled barcode schemes of 20 genes (Supplementary Table 2). The genes had a range of average copy numbers (from 1 to 10 copies per cell). Some genes (Cmk2, Npt1) are expressed at levels comparable or higher than the abundant tubulin subunits. The 12 remaining genes with low copy numbers (~1 per cell) were not measured since they are within the noise floor of our measurements. The two measurements agreed with an R^2=0.88 (*n*=1871 for the standard code, *n*= 1523 for the switched code. Fig. 3c, Supplementary Table 2, Supplementary Figure 12) in 19 of the 20 compared genes. As switching scrambles the barcode assignments for genes with different expression levels, the consistent results in this experiment showed that barcodes of the high copy number genes did not leak into the low copy number barcodes and bias their readings.

Combinatorial regulation in the calcium response transcriptional network

Cells often need turn on the expression of a large battery of genes in response to external stress. We analyzed the expression of a battery of genes in the calcium stress regulon that respond to pulses of Crz1 and Msn2. In Fig. 4, each column represents a single yeast cell and the mRNA abundances measured for each gene under 50 mM Ca^{2+} stress. From this data it is clear that the transcriptional response to calcium varies widely among individual cells.

To characterize the heterogeneity in regulon expression in single yeast cells, we first examined the variability in the abundances of each mRNA species. The distribution of expression levels for most genes included a low basal state and a long-tailed high-expression mode corresponding to transcriptional bursts (Supplementary Fig. 13), similar to those observed in previous smFISH experiments in yeast[25]. Since these transcriptional bursts are triggered by a TF pulse in our experiment, we asked whether our target genes burst synchronously with each other. Previous two-color smFISH experiments[26] in yeast found that targets of a constitutively active TF in the galatcose response network as well as metabolic genes were well correlated (R^2=0.5-0.7). We found a range of correlation coefficients from −0.11 to 0.78 between pairs of calcium responsive genes (Supplementary Fig. 14,15), indicative of heterogeneous coordination throughout the regulon.

Given this heterogeneous regulon expression profile, we next asked whether there were subgroups of genes that tend to burst together and whether they correspond to different regulatory architectures. We resolved two distinct clusters for the Crz1 responsive genes (*P*=0.09, 0.08 respectively) (Fig. 4, Supplementary Fig. 16). In the first cluster, Crz1 genes were preferentially clustered based on their coexpression with Msn2 genes. The promoter sequences of many of these genes, such as *PUN1*, *YLR194C*, *RCN2*, *NPT1* contained Msn2

binding sites along with Crz1 binding sites. Genes in the other cluster predominantly contained only Crz1 binding sites, with the exception of *YPS1* and *PMC1*, which also contain Msn2 sites. The lower average correlations between pairs of Crz1 and combinatorially regulated genes (R^2 = 0.36, *SD* = 0.16, *n* = 62 cells) compared to those among pure Crz1 targets (R^2 = 0.55, *SD* = 0.17, *n*=62) or among the combinatorial targets (R^2 = 0.52, *SD* = 0.11, *n* = 62 cells) indicated different sources of inputs on the target genes.

Since Crz1 and Msn2 pulses are not synchronized in individual cells[23] at the low calcium (50 mM) levels used in this experiment, we hypothesized that the two clusters of cells in the data corresponds to cells responding to either a Crz1 pulse or a Msn2 pulse (Fig. 4, Supplementary Note). We tested this by knocking out the Crz1 and Msn2 pathway separately and visualizing the loss of the associated expression clusters (Supplementary Figs. 17-20). In cells treated with FK506, an inhibitor of the Crz1 phosphotase Calcineurin, the overall expression level was lower (Fig. 5a). Conversely, in the Msn2– Msn4– cells, all targets genes should be expressed in response only to Crz1 pulses **(Supplementary Note)**. This is consistent with the observed coordination between pure Crz1 and combinatorial targets (Fig. 5e), as well as the disappearance of the cell cluster with only combinatorial genes expressed (Supplementary Fig. 18). This experiment indicates a modulatory effect for Msn2 on Crz1 target genes.

These experiments elucidated the combinatorial role of Msn2 and Crz1 on the stress response network and showed that single cell transcriptional profiling can be a valuable tool for identifying combinatorial control architecture in regulatory networks.

Discussion

Under a super-resolution microscope, cells can become virtual microarrays with giga-pixel information density. We have demonstrated two barcoding methodologies, spectral and spatial, and shown that by using 7 of the current super-resolution fluorophore pairs, transcripts from 32 genes can be detected simultaneously in single yeast cells using spectral barcoding.

Super-resolution barcoding can be dramatically scaled up with the techniques presented in this paper. We found that the typical transcript size is about 100 nm (Figs. 1 and 2,) and that transcripts are uniformly random distributed in yeast cells (Supplementary Fig.21) at a mean density of 1.9±1.5 barcodes/μm^2 (*n*=2463, s.d.). There are local regions that contain multiple mRNA barcodes within a diffraction-limited spot (Supplementary Fig.9, Supplementary Note) necessitating the use of super-resolution imaging. However, globally, super-resolution barcodes of the 32 profiled genes occupy less than 2% of the 2D super-resolution space available in a yeast cell, leaving a large amount of space open for further multiplexing. With one of the additional available emitters[27], ${}_{12}C_5$=792 genes can be spectrally coded.

Spectral Barcoding has a reconstruction fidelity dependent only on the hybridization of fluorophores, does not require linearization, has a low photon requirement and can be hybridized in a distributed pattern to increase robustness (Table 1). These advantages make it an excellent technique for barcoding mRNA and other molecules with unknown structures

with the existing SRM dye palette. In comparison, spatial barcoding has a localized hybridization pattern, a lower reconstruction fidelity due to the much higher photon requirements and the tertiary structure of the barcoded molecule **(Table 1)**. It has the advantage of scaling and may be more applicable to chromosome and splice isoform barcoding.

Scaling up beyond 1,000 genes will require the implementation of 3D SRM[28] to improve the axial resolution by a factor of 20, and the synthesis of an additional far-IR fluorophore as an emitter ($_{18}C_6$=18,532 genes can be spectrally coded with 1 additional emitter). At this higher barcode density, more sophisticated computational algorithms will also be necessary to identify barcodes in an intelligent and automatic fashion (Supplementary Note).

Super resolution barcoding provides several advantages over existing transcriptional profiling techniques. First, direct imaging of the sample preserves the spatial information both within cells and amongst cells. With the application of light sheet microscopy[29], this technique can be extended into optically thick samples without photobleaching associated with z-sectioning in epi-fluorescence microscopy. This advantage makes it a powerful tool in studying signaling in heterogeneous systems such as microbial ecosystems, tissue and embryos, where interactions among different cellular populations play an essential role in cellular decisions. Second, because of the single-molecule and *in situ* nature of the technique, the method is quantitative and avoids intrinsic bias in RNA extraction and conversion to cDNA. Lastly, many cells can be imaged simultaneously under a microscope quickly and throughput can be scaled up without considerable costs, compared to the high cost and long waiting time for sequencing single cells. After the initial cost of the probe set synthesis, the probe set can be hybridized many thousands of times to wild-type and mutant organisms. Super-resolution barcoding even at its current throughput can be a useful follow-up to existing high-throughput transcriptomics techniques by allowing genes of interest to be monitored with single cell resolution in spatially complex samples.

Lastly, the combinatorial labeling scheme can be applied to many types of molecules *in situ*. It is a short leap to consider combinatorial labeling of chromosomes and proteins[30], for single cell proteomics and ChIP experiments. We hypothesize that for many types of biochemical techniques, such as microarrays, there is an equivalent *in situ* single-cell experiment possible through super-resolution barcoding, removing the need for spatial separation traditionally performed by gels or dilution on a chip.

Methods

Probes design, Purification and Hybridization

25mer oligonucleotide probes (Biosearch Inc) were designed to match melting temperature whenever possible (Supplementary Table 3). FISH probes were designed with 2 base pair spacing between probes to allow efficient reactivation of the pair dyes, often leading to varying melting temperature between probes. Alexa Fluor 405, Alexa Fluor 488, and Cy3 were used as the activators and Cy5, Alexa Fluor 680, and Alexa Fluor 750 (Invitrogen) as the switchable dye. Labeling and purification of the probes followed the protocol in reference 16. Yeast cells were grown in YPD with 50 mM $CaCl_2$ and fixed in log growth

phase. Cells were also treated with 0.1% $NaBH_4$ for 30 minutes before the ethanol permeabilization step. We found the $NaBH_4$ treatment decreased the auto-fluorescent background of fixed yeast cells. Cells were stored at −20 °C in Eppendorf tubes and aliquoted out for hybridization experiments. Cells were hybridized with the probes overnight at room temperature in 20% formamide and 10% dextran sulfate. For all smFISH experiments, 12 probes (Supplementary Table 3) were hybridized to mRNA. For super resolution barcoding, sets of between 4 and 5, 6 to 7 probe pairs were used for Cy5, A680 and A750 respectively. After hybridization, cells were washed in 10% formamide and .2X SSC solution 3 times and imaged.

Imaging

For conventional epi-fluorescence microscopy, images were acquired on an Olympus IX81 with a 100x sapo objective with laser illumination at 532 nm, 594 nm, and 640 nm. Images were acquired with Micromanager software and an Andor Ikon-M DU934 BV CCD. Conventional fluorescence images were acquired in 3 different fluorescence channels (Semrock zero line filters). The centroids of the diffraction-limited FISH spots were calculated in each channel and the images were aligned by center of mass alignment of co-localized fluorescent spots between channels. This was sufficient for alignment without correcting for rotation and dilation.

Super-resolution imaging was performed on a Nikon TI-eclipse microscope with PFS autofocus lock. Standard super-resolution imaging buffers with glucose oxidase and β-mercaptoethanol were used[13]. The imaging lasers, a 640 nm laser along with a 30 mW 691 nm and a 30 mW 730 nm laser (Coherent Lasers) were brought to the sample through a 100x TIRFM objective. 405 nm, 473 and 556 lasers were used as activation lasers and imaging automation was controlled by Micromanager software. Two illumination pathways were used on the microscope, an imaging pathway containing the 640, 691 and 730 nm lasers and an activation pathway consisting of the remaining lasers. The activation pathway contained a automated filter wheel (Thorlabs) with ND filters .3, .6, 1 and 1.6 and an Uniblitz shutter, enabling automatic control of activation power by a custom written Micromanager script. No excitation filters were used in the microscope, enabling both activation and imaging wavelengths to reach the sample. Exposures times for all STORM movies were 150 ms for Cy5, 400 ms for Cy5.5 and 300 ms for Cy7.

For all STORM experiments, the imaging routine was designed to minimize crosstalk and photobleaching. To reduce activator crosstalk, every imaging routine sequentially used activation lasers from 556 nm, to 473 nm to 405 nm. As the absorption spectra of the activator dyes slightly overlap, this allowed us to minimize cross-activation by preferentially bleaching the overlapping dye before the routine reached the adjacent activation laser.

For the spatial coding images, samples were first imaged with only the 640 nm laser for 100 frames to switch off Cy5 and to determine the non-specific blinking rate. Then, 100 frames were acquired in each activation channel by co-illumination of the activation and the imaging lasers. This reduced the crosstalk among the different activation channels.

For the spectral coding images, the samples were first bleached in all imaging channels for 6 frames. Samples were imaged in order of the activators, starting at 556 nm to 473 nm and to 405 nm. For the 556 nm laser, the microscope first imaged the 730 nm followed by either the 640 nm or 691 nm laser for 48 cycles. This pattern minimized the Cy5 and A680 signal loss due to bleaching of both dyes by the 640 nm and 691nm lasers. For the remaining activation lasers, the microscope cyclically imaged the 730 nm laser followed by the 640 nm laser. Samples are only illuminated with the activation light in the A750 channel. This imaging pattern minimized photobleaching of A750 while still allowing us to observe adequate activation of the A680 and Cy5 probe pairs. Throughout the imaging routine, for every 4 cycles of the imaging lasers, 2 cycles without the activator lasers were acquired. This enabled us to rule out much of the false positive and nonspecific blinking events in the images. Activation powers were selected to maximize activation rate while avoiding crosstalk amongst the activators. Following the imaging cycle with all emitters, the 640 nm laser was continuously used to image Cy5 for another 30 frames of specific activation with an ND filter added to the pathway to bring the activation laser power to 50% of the original activation. The acquisition sequence is illustrated in Supplementary Fig. 11.

The activation lasers were controlled by an Arudino microcontroller board and a servo motor shutter. Fluorescent beads (Invitrogen F-8810) were used as fiducial markers to correct for stage drifts. The microscope stages (Prior and ASI) were automated and controlled by acquisition software to enable multi-position imaging.

Analysis

Images from diffraction limited and super-resolution experiments were analyzed with a custom written Mathematica script, available upon request. In the analysis, the beads were first aligned to determine the stage drifts. Beads emit on the order of 50,000 photons per image, and could be localized to a few nanometers. Beads close to the cells were eliminated from analysis as the switching of fluorophores in cells can disrupt the bead alignment. Then, fluorophores were selected from each image by intensity thresholding (600 photons for cy5, 400 photons for A680 and 250 photons for A750), and their centroid calculated by Gaussian fitting. We did not reject activation events that involved multiple fluorophores in the analysis since they came predominantly from single clusters and allowed us to preserve more of the photons collected for the analysis. In principle, multiple activation frames can be compared to neighboring frames to extract localization information.

To determine which of the barcode colors are present instead of nonspecific activation, we compared the number of activation events occurring in the specific activation channel versus the activations observed in frames with no activation. If the frequency of nonspecific activation events exceeded that of specific activation events, then that channel was rejected. As several of the activators and emitter channels can crosstalk into other channels, we quantified the crosstalk ratio and rejected activation events if they fell below our thresholds (further discussion in the Supplementary Note). Switching events that spatially cluster together were grouped to display the resolved barcode. For spectral barcodes, activations were clustered together on a 184 nm grid for simplicity since most RNAs are contained within a diameter of 100 nm. Activation events near grid vertices were assigned to a

neighboring region containing localizations of the same fluorophore pair. For 3 color spatial barcodes, the center color was determined by finding the position that was not one of the two localized colors separated by the longest distance. Cell positions were determined by manual segmentation. Barcodes were collected and tabulated for each single cell. Cross-correlation was calculated using the correlation function in Mathematica and the standard errors were calculated from re-sampling the data 100 times.

QPCR Measurements

For the RT-PCR experiments, *S. cerevisiae* were grown in YPD media and were induced with 50mM $CaCl_2$ for two hours. The cells were then spheroplasted with zymolyase, RNase inhibitor and BME. The RNA isolated using the Trizol Plus RNA Purification System (Invitrogen). The extract was treated with recombinant Dnase I (Roche) to digest genomic DNA. Reverse transcription of the RNA was performed using the iScript Reverse Transcription Supermix (Bio-Rad). Aliquots of the resulting cDNA were diluted 1:10 and 1:100 with water. Quantitative PCR was performed using Maxima Sybr Green and Flourescein qPCR Master Mix (Fermentas) on a Bio-Rad CFX machine. Quantitative PCR was performed on each gene with the undiluted cDNA, 1:10 dilution cDNA and 1:100 dilution cDNA, with each dilution repeated in triplicate (Figure 3). The PCR efficiency for each gene was calculated from a linear fit of the three dilution points. The abundance of the RNA was determined from the amplification efficiency. Primer sequences are presented in Supplementary Table 4.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We credit B. Wold with discussions that led to this work. We thank M. Elowitz for lending space and equipment in his laboratory. We thank T. Zhiyentayev, H.Q. Li, and X. Wang for assistance with experiments. We are grateful to A. Raj for technical assistance with FISH, and X.W. Zhuang and her group for STORM. We thank A. Eldar, S. Fraser, G.W. Li, J. Levine, J. Locke for discussion and reading of the manuscript. This work was supported by a Beckman Institute seed grant and a NIH New Innovator Award 1DP2OD008530.

We dedicate this paper to Xiaoliang Sunney Xie on the occasion of his 50th birthday.

References

1. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007; 316:1497–502. [PubMed: 17540862]
2. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods. 2008; 5:621–8. [PubMed: 18516045]
3. Nagalakshmi U, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008; 320:1344–9. [PubMed: 18451266]
4. Schena M, Shalon D, Davis RW, Brown PO. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. Science. 1995; 270:467–470. [PubMed: 7569999]
5. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature. 2000; 403:335–8. [PubMed: 10659856]
6. Chalfie M, Tu Y, Euskirchen G, Ward W, Prasher D. Green fluorescent protein as a marker for gene expression. Science. 1994; 263:802–805. [PubMed: 8303295]

7. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005; 123:1025–36. [PubMed: 16360033]
8. Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. Nature. 2006; 440:358–62. [PubMed: 16541077]
9. Yu J, Xiao J, Ren X, Lao K, Xie XS. Probing gene expression in live cells, one protein molecule at a time. Science. 2006; 311:1600–3. [PubMed: 16543458]
10. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS biology. 2006; 4:e309. [PubMed: 17048983]
11. Hell SW. Far-field optical nanoscopy. Science. 2007; 316:1153–8. [PubMed: 17525330]
12. Betzig E, et al. Imaging intracellular fluorescent proteins at nanometer resolution. Science. 2006; 313:1642–5. [PubMed: 16902090]
13. Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nature methods. 2006; 3:793–796. [PubMed: 16896339]
14. Hess ST, Girirajan TPK, Mason MD. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophysical journal. 2006; 91:4258–72. [PubMed: 16980368]
15. Femino AM. Visualization of Single RNA Transcripts in Situ. Science. 1998; 280:585–590. [PubMed: 9554849]
16. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods. 2008; 5:877–879. [PubMed: 18806792]
17. Lowenstein MG. Long-Range Interphase Chromosome Organization in Drosophila: A Study Using Color Barcoded Fluorescence In Situ Hybridization and Structural Clustering Analysis. Molecular Biology of the Cell. 2004; 15:5678–5692. [PubMed: 15371546]
18. Levsky JM, Shenoy SM, Pezo RC, Singer RH. Single-cell gene expression profiling. Science. 2002; 297:836–40. [PubMed: 12161654]
19. Thompson RE, Larson DR, Webb WW. Precise nanometer localization analysis for individual fluorescent probes. Biophysical journal. 2002; 82:2775–83. [PubMed: 11964263]
20. Yildiz A, et al. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. Science. 2003; 300:2061–5. [PubMed: 12791999]
21. Bates M, Dempsey GT, Chen KH, Zhuang X. Multicolor super-resolution fluorescence imaging via multi parameter detection. Chemphyschem. 2012; 13:99–107. [PubMed: 22213647]
22. Barish RD, Schulman R, Rothemund PWK, Winfree E. An information-bearing seed for nucleating algorithmic self-assembly. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:6054–9. [PubMed: 19321429]
23. Cai L, Dalal CK, Elowitz MB. Frequency-modulated nuclear localization bursts coordinate gene regulation. Nature. 2008; 455:485–90. [PubMed: 18818649]
24. Gasch AP, et al. Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Mol. Biol. Cell. 2000; 11:4241–4257. [PubMed: 11102521]
25. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. Nature structural & molecular biology. 2008; 15:1263–71.
26. Gandhi SJ, Zenklusen D, Lionnet T, Singer RH. Transcription of functionally related constitutive genes is not coordinated. Nature structural & molecular biology. 2011; 18:27–34.
27. Dempsey GT, Vaughan JC, Chen KH, Bates M, Zhuang X. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. Nature methods. 2011; 8:1027–36. [PubMed: 22056676]
28. Huang B, Wang W, Bates M, Zhuang X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. Science. 2008; 319:810–3. [PubMed: 18174397]
29. Cella Zanacchi F, et al. Live-cell 3D super-resolution imaging in thick biological samples. Nature methods. 2011; 8:1047–9. [PubMed: 21983925]
30. Agnew HD, et al. Iterative in situ click chemistry creates antibody-like protein-capture agents. Angewandte Chemie (International ed. in English). 2009; 48:4944–8. [PubMed: 19301344]

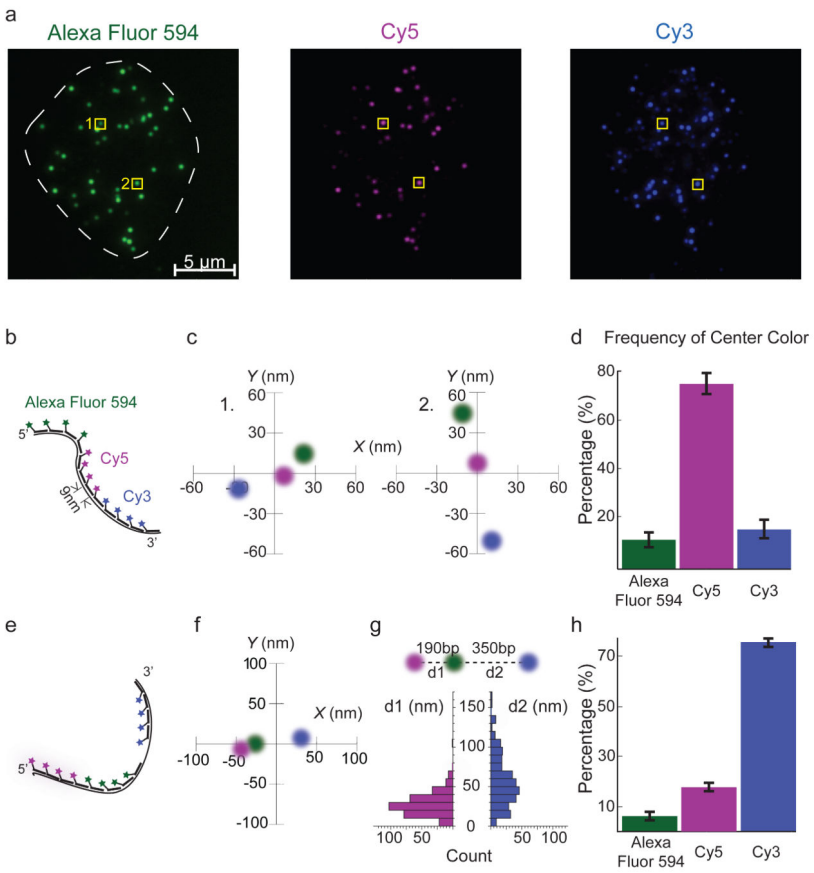

Figure 1.

Spatial ordering of fluorophores on mRNAs can be resolved by gaussian centroid localization. (**a**). Fluorescence images of *PUN1* probes hybridized in a single budding yeast cell, shown in each channel. (**b**). Schematic of labeled 25mer oligonucleotides hybridized to *PUN1* mRNA. (**c**). Reconstructions of the centroids of spots 1 and 2 following localization by Gaussian fitting and image alignment. (**d**). The percentage of co-localized *PUN1* three-color dots that can be reconstructed in the above image (**a**) with the correct barcode (*n*=28, Correct Order=74±8%, s.e.m.). (**e**). Schematic of probe-set hybridized to GFP mRNA with different order and distances between the probes positions. **(f).** Gaussian fitting reconstruction of this probe set. **(g).** The distance between the resolved centroids positions (*d1*=27.93±14 nm, *d2*= 56±33 nm s.d.) is proportional to the intramolecular distance between barcode positions (190 and 350bp). **(h).** The frequency of barcode identification for this probe set (*n*=327, correct order=76±2%, s.e.m.).

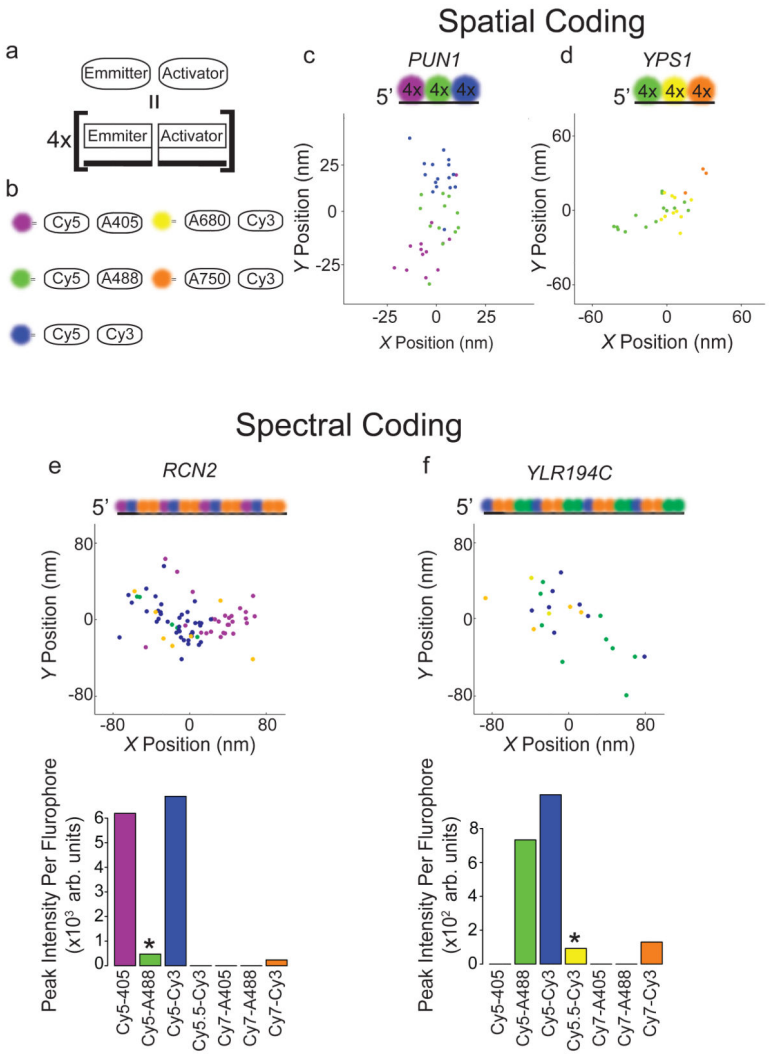

Figure 2.

Super-resolution imaging enables combinatorial labeling of individual transcripts. **(a).** Schematic of super-resolution barcoding scheme. For each super-resolution probe pair, four acceptor-emitter pairs are hybridized in sequence for redundancy. **(b).** Each barcode color consists of an activator (Alexa Fluor 405, Alexa Fluor 488, and Cy3) labeled oligo adjacent to an emitter (Cy5, Alexa Fluor 680 and Alexa Fluor 750) labeled oligo. **(c).** *PUN1* mRNA 3 position spatial barcode. The order of the probes is shown schematically in the cartoon. A localization scatterplot in which each dot represents an activation of a fluorophore pair is shown. (**d**). *YPS1* mRNA 3 position spatial barcode with 3 different emitters. **(e).** *RCN2* mRNA spectral 3 Position Barcode. Probe Positions are scattered throughout the mRNA, enabling robust hybridization. Bar height in histogram is equal to the number of integrated peak pixel counts detected for each fluorophore pair over time. Cy5-A405, Cy5-A488, Cy5-Cy3 and A750-Cy3 are detected with 6195, 471, 6881 and 235 integrated counts respectively. Cy5-A488 (green and asterisk) is present due to cross-talk from Cy5-Cy3 (blue), and rejected based on the threshold measurements in Supplementary Figs. 10. Note that the A750 based dye pairs give fewer photons than Cy5 dye pairs, but are readily

detected with less crosstalk. **(f).** YLR194c mRNA spectral 3 position barcode. Cy5-A488, Cy5-Cy3, A750-Cy3 and A680-Cy3 are detected with 773, 999, 130 and 92 integrated peak counts. A680-Cy3 (yellow and asterisk) was rejected due to crosstalk from Cy5-Cy3.

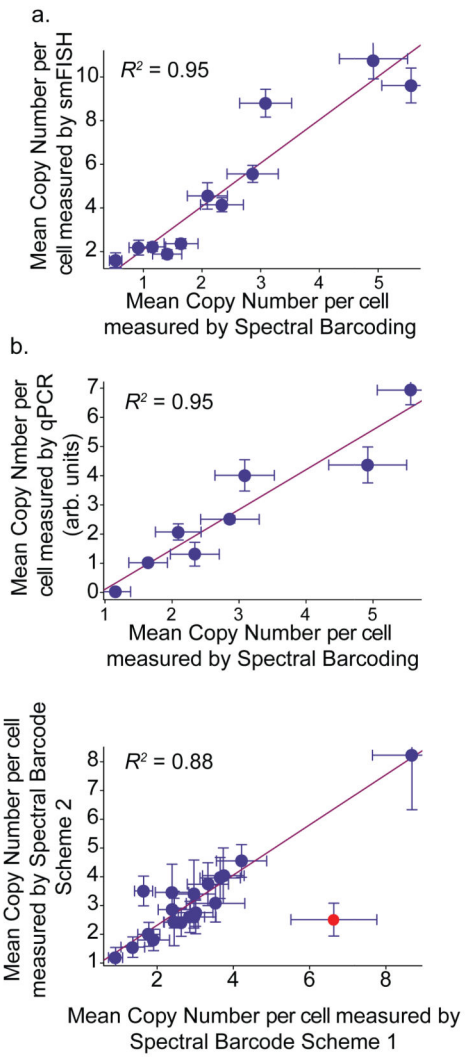

Figure 3.

Validation of mRNA quantitation by super-resolution barcoding. (**a**). Comparison of smFISH results with super-resolution barcodes gives an R^2 = 0.95, with a slope of 2.05. 11 genes were FISHed, including 8 crz1 specific genes, 1 Msn2 target gene, and 2 aging and stress genes. (**b**). For the qPCR experiment, 8 Crz1 genes were quantified. **(c).** Robustness of mRNA quantitation measured by two different barcode schemes. Mean copy-number measurements for barcoding schemes are displayed as points in the scatterplot. A regression with an R^2 value of 0.88 was obtained following removal of the one outlier connoted by a red circle. The outlier was removed due to its high cook's distance of 2.1 (Supplementary Fig. 12).

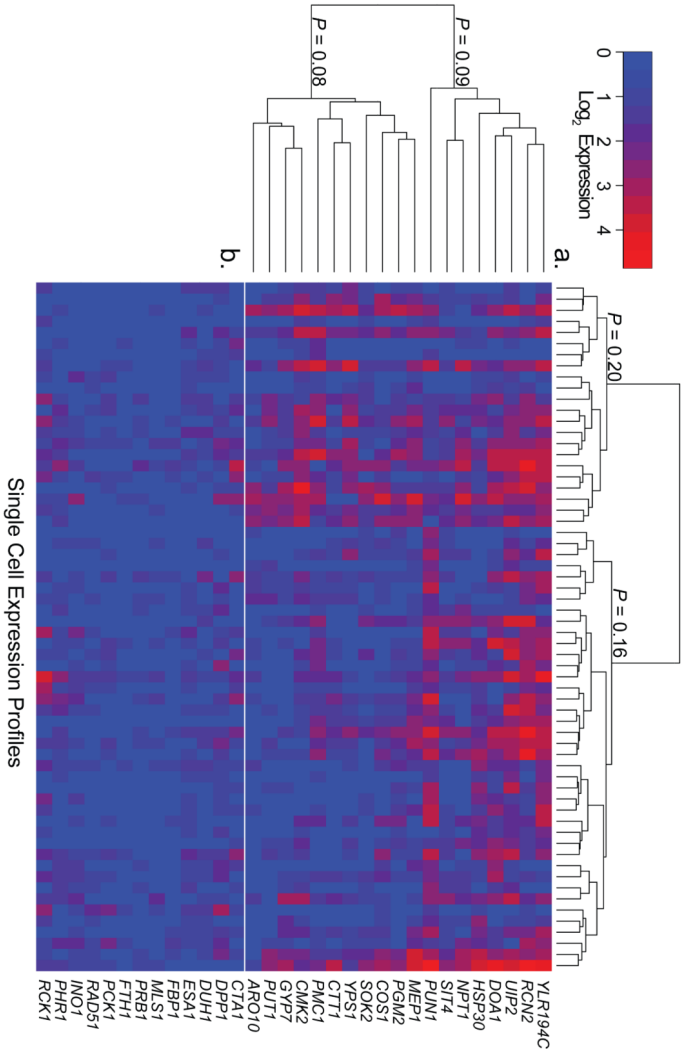

Figure 4.

Single cell expression profiles of 32 mRNAs. Each column corresponds to the expression profile of a single yeast cell. Cells, and genes in (**a**) are clustered using agglomerative hierarchal clustering on the correlation between species using Ward's criterion. The *P*-values for the secondary clusters are given by bootstrap resampling and placed adjacent to these clusters. **(a).** Genes responsive to Crz1 and Msn2. Genes can be broadly clustered into two classes, one largely containing genes regulated by both Crz1 and Msn2 (*P*=0.09, upper cluster) and one largely containing genes regulated by Crz1 (*P*=0.08, lower cluster). Cells are grouped in two distinct clusters, one showing correlations amongst the expression of all genes regulated by Crz1 (*P*=0.2, left cluster), the other with large expression correlations amongst combinatorial genes (*P*=0.16, right clusters). **(b).** Additional measured genes are shown.

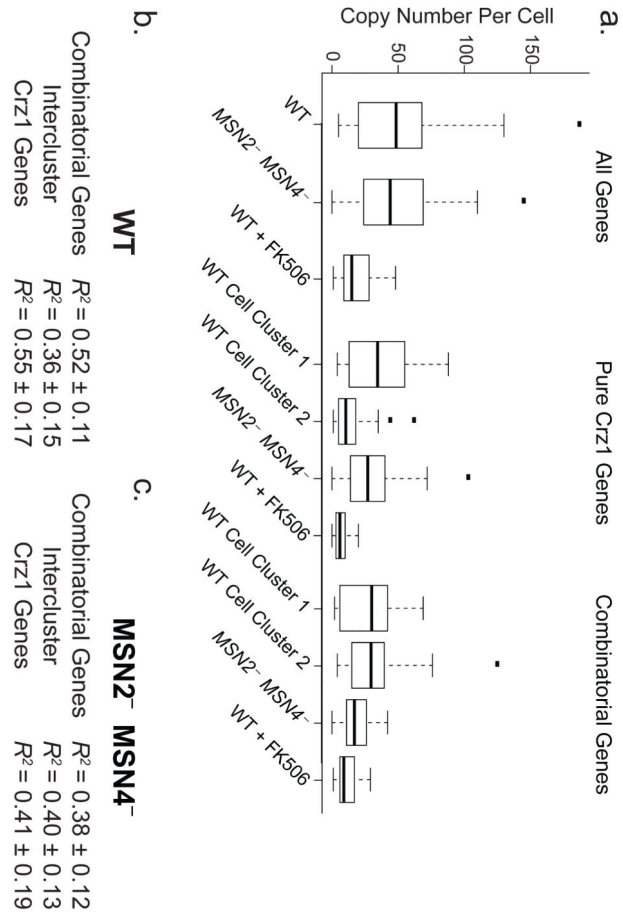

Figure 5.

Msn2 and Crz1 combinatorially affect target regulons. All data in this analysis comes from the genes in Fig. 4a, genes displayed in Fig. 4b are not included in these analyses. **(a).** Total number of regulon-specific mRNAs in single yeast cells shown in box plots. Pure Crz1 targets are expressed in WT and in the Msn2– and Msn4– cells, but are repressed in FK506 treated cells. WT cells are clustered into two groups (Fig. 4a), WT Cell Cluster 1 and WT Cell Cluster 2, corresponding to cells with Crz1 or Msn2 pulses respectively. Combinatorial targets are repressed in the *Msn2–* and *Msn4–* as well as in the FK506 treated cells. **(b-c).** Averaged correlation coefficients between pairs of genes in the combinatorial or pure Crz1 target clusters. The detailed pairwise gene expression scatter plots are shown in Supplementary Figs. 14,15 .

Table 1

Comparative advantages of SRM barcoding techniques.

Barcode Type	Hybridization Pattern	Spatial Reconstruction Fidelity	Resolution Requirement	Minimum required Flurophore Emission	Linearization required	Multiplex Scaling
Spectral	distributed	100%	100 nm	~400 photons	No	$p!/(p-n)!/n!$
Spatial	localized	74%	20 nm	~3000 photons	Yes	$p!/(p-n)!/2$

*Calculations of multiplex scaling use *p* as the number of fluorophores and *n* as the number of positions.